Difference-in-Difference: A novel approach

Will Zhang

University of Washington

June 7, 2023

Table of contents

Background

Richardson and Tchetgen Tchetgen (2021) (Richardson *et al.*, 2023) introduced Bespoke IV (generalized diff-in-diff or GDID), which offer an alternative approach to identify causal effects under different assumptions than what's required for standard DID.

Shortcomings of existing estimators:

- DID requires treatment parallel trend assumption for identification and exogenous policy, which do not always hold.
- GDID requires an alternative set of assumptions including instrument parallel trend assumptions, exhibits lower efficiency in terms of standard error, etc.

Intuition

Given different sets of assumptions, we can fit

1. OLS (fitting a linear regression of $Y_i(t)$ on a combination of covariates) to derive causal effect of interest in DID
2. 2SLS (using valid covariates as instruments) to derive causal effect of interest in GDID

However, assumptions may not hold for either procedure. I propose a penalized OLS regression with 2SLS as penalization. Doing such, the estimator exhibits

- a set of possible causal parameters, where the result depends on the tuning parameter in penalized regression;
- potential distributional robustness properties given certain assumptions hold.

Notation

Below notations are described for our interest of average effect of treatment on the treated (ATT): $\mathrm{E}\left[Y_i^1(t_1) - Y_i^0(t_1) \mid A_i = 1\right]$

- A_i denote treatment, where $A_i = 1$ if individual i is treated, $A_i = 0$ otherwise.
- $Y_i(t)$ denotes the outcome of interest at time t, where a population is observed in two periods: a pre-treatment period, $t = t_0$; and, a post-treatment period, $t = t_1$.
- $Z_i(t)$ and $U_i(t)$ denote vectors of measured and unmeasured variables: $(Z_{i,1}, \ldots, Z_{i,p})^\top$, and similarly $(U_{i,1}, \ldots, U_{i,q})^\top$.
- $Y_i^a(t)$ denote individual's potential outcome at time t.

Proposed estimator: Robust diff-in-diff (RDID)

$$\theta_{\mathrm{RDID}}(\gamma) := \arg\min_{\theta} \{\mathcal{L}_{\mathrm{OLS}}(\theta) + \gamma \mathcal{L}_{\mathrm{IV}}(\theta)\}$$

where θ is the treatment effect, $\gamma \in [-1, \infty)$ is the tuning parameter, with defined loss functions:

$$\mathcal{L}_{\mathrm{OLS}}(\theta; Y, Z, A) := \mathrm{E}\left[Y_i(t) - \theta A_i t - \mu_i - \nu_t - \beta^\top Z_{it}\right]^2,$$

$$\mathcal{L}_{\mathrm{IV}}(\theta; Y, Z, A) := \mathrm{E}\left[Z_i\left(Y_{*i} - \alpha - \theta A_i\right)\right]^\top \mathrm{E}\left[Z_i Z_i^\top\right]^{-1} \mathrm{E}\left[Z_i\left(Y_{*i} - \alpha - \theta A_i\right)\right]$$

where we assume Z is not time-varying for simplicity.

- μ_i, ν_t denote individual and time fixed effect, respectively
- $Y_{*i} = Y_i(t_1) - Y_i(t_0)$
- $\alpha = \mathrm{E}[Y_i^{a=0}(t_1) - Y_i^{a=0}(t_0) \mid Z = 0]$ accounts for baseline difference of potential outcome of no treatment between two periods.

RDID as Anchor Regression (AR)

We can drop (i, t) notations for simplicity.

$$\mathcal{L}_{\mathrm{OLS}}(\theta; Y, A) := \mathrm{E}\left[Y_* - \alpha - \theta A\right]^2, \text{where } \alpha = \nu_{t=1} - \nu_{t=0}$$

$$\mathcal{L}_{\mathrm{IV}}(\theta; Y, Z, A) := \mathrm{E}\left[Z\left(Y_* - \alpha - \theta A\right)\right]^\top \mathrm{E}\left[ZZ^\top\right]^{-1} \mathrm{E}\left[Z\left(Y_* - \alpha - \theta A\right)\right]$$

Let $\Theta = (\alpha, \theta)^\top$ and $A_* = (1, A)^\top$, we express the penalized regression as Anchor Regression (AR):

$$\Theta_{\mathrm{RDID}}(\gamma) = \arg\min_{\Theta}\{\mathrm{E}[(Y_* - \Theta^\top A_*)^2] + \gamma \mathrm{E}[(P_Z(Y_* - \Theta^\top A_*))^2]\},$$

where P_Z denotes the L2-projection on the linear span from Z.

Closed form solution

Let $(\mathbf{Y}, \mathbf{Z}, \mathbf{A})$ consist of n row-wise random vector (Y, Z, A) respectively. We express sample analog

$$\hat{\Theta}^n_{\mathrm{RDID}}(\gamma) = \arg\min_{\Theta}\{\|\mathbf{Y}_* - \mathbf{A}_*\Theta\|_2^2 + \gamma\,\|\Pi_{\mathbf{Z}}(\mathbf{Y}_* - \mathbf{A}_*\Theta)\|_2^2\},$$

where $\Pi_{\mathbf{Z}} \in \mathbb{R}^{n\times n}$ is projection matrix. i.e. $\Pi_{\mathbf{Z}} := \mathbf{Z}\left(\mathbf{Z}^\top\mathbf{Z}\right)^{-1}\mathbf{Z}^\top$, we obtain the closed form solution

$$\hat{\Theta}^n_{\mathrm{RDID}}(\gamma) = \left[\mathbf{A}_*{}^\top\left(I + \gamma\Pi_{\mathbf{Z}}\right)\mathbf{A}_*\right]^{-1}\mathbf{A}_*{}^\top\left(I + \gamma\Pi_{\mathbf{Z}}\right)\mathbf{Y}_*,$$

where $\Pi_Z = \mathbf{Z}\left(\mathbf{Z}^\top\mathbf{Z}\right)^{-1}\mathbf{Z}^\top$ is the orthogonal projection onto the column space of $\mathbf{Z}$.

Assumptions for RDID

Suppose that $Z = (Z_1, Z_2)$ and instead of taking all of Z as a candidate bespoke instrumental variable:

- take Z_1 only as a bespoke instrumental variable
- Z_2 are additional measured covariates that we adjust for.

We begin by assuming:

1. Consistency for the treated, $Y(t_1) = Y^a(t_1)$, if $A = a$
2. Positivity (i.e., a small constant $\mathrm{c} > 0$, such that for any z such that $\Pr(Z = z \mid A = 1) > \mathrm{c}$ it must be that $\Pr(Z = z \mid A = 0) > \mathrm{c}$
3. No anticipation of future treatment (i.e., at t_0 individuals do not anticipate the treatment received at t_1), such that $\mathrm{E}[Y(t_0) \mid Z] = \mathrm{E}[Y^a(t_0) \mid Z]$), for all a.

Above are standard assumptions required for DID besides parallel trend assumption.

Additional Assumptions for RDID

4. Z_1 is relevant for predicting treatment: $\mathrm{E}[A \mid Z_1, Z_2]$ depends on Z_1;
5. No interaction between A and Z_1 in causing $Y^a(t_1)$ conditional on Z_2 and $A = 1$, such that
$\mathrm{E}\left[Y^a(t_1) - Y^0(t_1) \mid A = a, Z_1, Z_2\right] = \mathrm{E}\left[Y^a(t_1) - Y^0(t_1) \mid A = a, Z_1 = 0, Z_2\right]$
6. $\exists \gamma \in (0, \infty)$ s.t. $(\mathcal{C}_1)_{2,2} \times \text{bias}(\theta_{\mathrm{OLS}}) + (\mathcal{C}_2)_{2,2} \times \text{bias}(\theta_{\mathrm{IV}}) = 0$, where

$$\text{bias}(\theta_{\mathrm{OLS}}) = \mathrm{E}\left[Y^0(t_1) - Y^0(t_0) \mid A = 1\right] - \mathrm{E}\left[Y^0(t_1) - Y^0(t_0) \mid A = 0\right]$$

$$\text{bias}(\theta_{\mathrm{IV}}) = \frac{\mathrm{E}[Y^a(t_1) - Y^{t_0}(1) \mid A - a, Z_1, Z_2] - \mathrm{E}[Y^a(t_1) - Y^{t_0}(1) \mid A = a, Z_1 = 0, Z_2]}{\mathrm{E}[A \mid Z_1 = z_1, Z_2] - \mathrm{E}[A \mid Z_1 = 0, Z_2]}$$

$$\mathcal{C}_1 = I - \gamma \mathrm{E}[AA^\top]^{-1}\mathrm{E}[AZ^\top]\mathrm{E}[ZZ^\top]^{-1}\left(\mathrm{E}[ZZ^\top](I + \gamma \mathrm{E}[ZA^\top]\mathrm{E}[AA^\top]^{-1}\mathrm{E}[AZ^\top]\mathrm{E}[ZZ^\top]^{-1})\right)^{-1}\mathrm{E}[ZZ^\top]^{-1}\mathrm{E}[ZA^\top]$$

$$\mathcal{C}_2 = I - \frac{1}{\gamma}(\mathrm{E}[AZ^\top]\mathrm{E}[ZZ^\top]^{-1}\mathrm{E}[ZA^\top])^{-1}\mathrm{E}[AA^\top](\mathrm{E}[AA^\top](I + \frac{1}{\gamma}\mathrm{E}[AA^\top](\mathrm{E}[AZ^\top]\mathrm{E}[ZZ^\top]^{-1}\mathrm{E}[ZA^\top])^{-1}\mathrm{E}[AA^\top]))^{-1}\mathrm{E}[AA^\top]$$

RDID Identification

We want to identify the average treatment effect on treated

$$E\left[Y^{a=1}(t_1) - Y^{a=0}(t_1) \mid A = 1, Z\right] = \psi(Z)$$

Suppose assumptions 1-6 hold. For $z_1 \neq 0$ and $\gamma \in (0, \infty)$, we have

$$\psi(Z) = \frac{\theta_{\mathrm{RDID}} - (\mathcal{C}_1)_{2,1}\alpha_{\mathrm{OLS}} - (\mathcal{C}_2)_{2,1}\alpha_{\mathrm{IV}}}{(\mathcal{C}_1)_{2,2} + (\mathcal{C}_2)_{2,2}}$$

Structural equation model setup

We only assume the Z as the exogenous variable, and causal relations between Y, A, U, Z are ambiguous. Consider a possibly cyclic linear SEM, dropping $*$ notations, as in Rothenhäusler *et al.* (2018):

$$\begin{bmatrix} Y & A & U^\top \end{bmatrix} := \begin{bmatrix} Y & A & U^\top \end{bmatrix} B + Z^\top M + \varepsilon^\top,$$

- $Y \in \mathbb{R}$ is the difference in outcome between two periods,
- $A \in \mathbb{R}^2$ is the constant and treatment,
- $U \in \mathbb{R}^q$ are hidden variables with possible presence of causal relations,
- $Z \in \mathbb{R}^p$ are exogenous variables independent from the unobserved noise innovations ε.

Assuming $I - B$ is invertible, under intervention $\mathrm{do}(Z := v)$:

$$\begin{bmatrix} Y & A & U^\top \end{bmatrix} := (v^\top M + \varepsilon^\top)(I - B)^{-1},$$

Distributional robustness of RDID

Recall the solution to RDID empirical minimization problem:

$$\hat{\Theta}_{\mathrm{RDID}}(\gamma) = \left[\mathbf{A}^\top (I + \gamma \Pi_{\mathbf{Z}}) \mathbf{A}\right]^{-1} \mathbf{A}^\top (I + \gamma \Pi_{\mathbf{Z}}) \mathbf{Y},$$

RDID is well-identified even if the model is under-identified. Then, RDID aims to possess interventional robustness over causal inference.

$$\Theta_{\mathrm{RDID}}(\gamma) = \underset{\Theta \in \mathbb{R}^2}{\arg\min} \sup_{v \in C(\gamma)} E^{\mathrm{do}(Z:=v)} \left[\left(Y - \Theta^\top A\right)^2\right],$$

where $C(\gamma) := \left\{v : \Omega \to \mathbb{R}^q : \mathrm{Cov}(v, \varepsilon) = 0, \mathrm{E}\left[vv^\top\right] \preceq (\gamma + 1) E\left(ZZ^\top\right)\right\}$.

K-class estimator generalization

Define $\mathbf{X} = [\mathbf{Z}\mathbf{A}]$ with $\mathbf{Z} \in \mathbb{R}^{n\times((2+q))}$. Given matrices are well-defined assumptions hold, with parameter $\kappa \in \mathbb{R}$:

$$\hat{\Theta}^{n}_{\mathrm{RDID}}(\kappa) = \left[\mathbf{X}^{\top}\left(I - \kappa\Pi^{\perp}_{\mathbf{Z}}\right)\mathbf{X}\right]^{-1}\mathbf{X}^{\top}\left(I - \kappa\Pi^{\perp}_{\mathbf{Z}}\right)\mathbf{Y},$$

where $I - \kappa\Pi^{\perp}_{\mathbf{Z}} = I - \kappa\left(I - \Pi_{\mathbf{Z}}\right) = (1-\kappa)I + \kappa\Pi_{\mathbf{Z}}$.
K-class estimator with no included exogenous variables , coincides with the RDID(AR) estimator with penalty parameter $\gamma = \kappa/(1-\kappa)$, i.e.

- for $\kappa < 1$, $\hat{\Theta}^{n}_{\mathrm{K}}(\kappa) = \hat{\Theta}^{n}_{\mathrm{RDID}}\left(\frac{\kappa}{1-\kappa}\right)$;
- for $\gamma > -1$, $\hat{\Theta}^{n}_{\mathrm{RDID}}(\gamma) = \hat{\Theta}^{n}_{\mathrm{K}}(\gamma/(1+\gamma))$;

therefore, both exhibit distributional robustness properties under some assumptions, as shown in Jakobsen and Peters (2021).

RDID Robustness Properties

Causal inference versus minimizing prediction error

Work in progress.

D. B. Richardson and E. J. Tchetgen Tchetgen, *American Journal of Epidemiology*, 2021, **191**, 939–947.

D. B. Richardson, T. Ye and E. J. Tchetgen Tchetgen, *Epidemiology*, 2023, **34**, 167–174.

D. Rothenhäusler, N. Meinshausen, P. Bühlmann and J. Peters, *Anchor regression: heterogeneous data meets causality*, 2018, `https://arxiv.org/abs/1801.06229`.

M. E. Jakobsen and J. Peters, *The Econometrics Journal*, 2021, **25**, 404–432.